df.createOrReplaceTempView("df")

spark.sql("""

SELECT \*,

CASE

WHEN location\_description IN (

'SINGLE FAMILY DWELLING', 'MULTI-UNIT DWELLING', 'APARTMENT',

'CONDOMINIUM/TOWNHOUSE', 'HOUSE', 'MOBILE HOME/TRAILERS'

) THEN 'Residential'

WHEN location\_description IN (

'GAS STATION', 'JEWELRY STORE', 'GROCERY STORE', 'LIQUOR STORE',

'CLOTHING STORE', 'SUPERMARKET', 'BAR/COCKTAIL/NIGHTCLUB',

'NAIL SALON', 'PHOTO/COPY', 'RESTAURANT/FAST FOOD'

) THEN 'Commercial'

WHEN location\_description LIKE 'MTA%' OR location\_description IN (

'BUS STOP', 'SUBWAY PLATFORM', 'TRAIN TRACKS', 'PARKING LOT',

'TAXI', 'AIRPORT TERMINAL'

) THEN 'Transit'

WHEN location\_description IN (

'HOSPITAL', 'DOCTOR/DENTIST OFFICE', 'NURSING HOME',

'CLINIC', 'VETERINARIAN', 'HOSPICE', 'MEDICAL MARIJUANA'

) THEN 'Medical'

WHEN location\_description IN (

'ELEMENTARY SCHOOL', 'HIGH SCHOOL', 'COLLEGE',

'UNIVERSITY', 'PRIVATE SCHOOL', 'TRADE SCHOOL'

) THEN 'Educational'

WHEN location\_description IN (

'GOVERNMENT FACILITY', 'FIRE STATION', 'POLICE FACILITY',

'POST OFFICE', 'COURTHOUSE', 'JAIL/DETENTION CENTER'

) THEN 'Government'

WHEN location\_description IN (

'CHURCH', 'MOSQUE', 'SYNAGOGUE', 'TEMPLE',

'PLACE OF WORSHIP'

) THEN 'Religious'

WHEN location\_description IN (

'PARK/PLAYGROUND', 'STREET', 'ALLEY', 'BEACH', 'LAKE',

'RIVER', 'YARD', 'OPEN LOT'

) THEN 'Outdoor'

ELSE 'Other'

END AS location\_category

FROM df

""").show()  
  
  
# Step 1: Register temp view

df.createOrReplaceTempView("victim\_data")

# Step 2: Write Spark SQL query

result = spark.sql("""

SELECT \*,

CASE

WHEN victim\_race IN ('WHITE') THEN 'White'

WHEN victim\_race IN ('BLACK') THEN 'Black or African American'

WHEN victim\_race IN ('WHITE HISPANIC') THEN 'White Hispanic'

WHEN victim\_race IN ('BLACK HISPANIC') THEN 'Black Hispanic'

WHEN victim\_race IN ('ASIAN / PACIFIC ISLANDER') THEN 'Asian / Pacific Islander'

WHEN victim\_race IN ('AMERICAN INDIAN/ALASKAN NATIVE') THEN 'American Indian / Alaska Native'

WHEN victim\_race IN ('OTHER') THEN 'Other'

WHEN victim\_race IN ('UNKNOWN') THEN 'Unknown'

WHEN victim\_race IN ('W') THEN 'White'

WHEN victim\_race IN ('B') THEN 'Black or African American'

WHEN victim\_race IN ('A') THEN 'Asian'

WHEN victim\_race IN ('I') THEN 'American Indian / Alaska Native'

WHEN victim\_race IN ('H') THEN 'Hispanic'

WHEN victim\_race IN ('X') THEN 'Unknown'

WHEN victim\_race IN ('P', 'F', 'L', 'U', 'V', 'O', 'D', 'K', 'Z', 'C', 'S', 'J', 'G') THEN 'Other'

WHEN victim\_race IS NULL OR victim\_race = '-' OR victim\_race = '(null)' THEN 'Unknown'

ELSE 'Other'

END AS race\_group

FROM victim\_data

""")